

## Homework #3: Wrapper

Due: 1/30, Wednesday

100 points

(TA handling this homework: Binh Vu [binhlvu@usc.edu](mailto:binhlvu@usc.edu))

### Description:

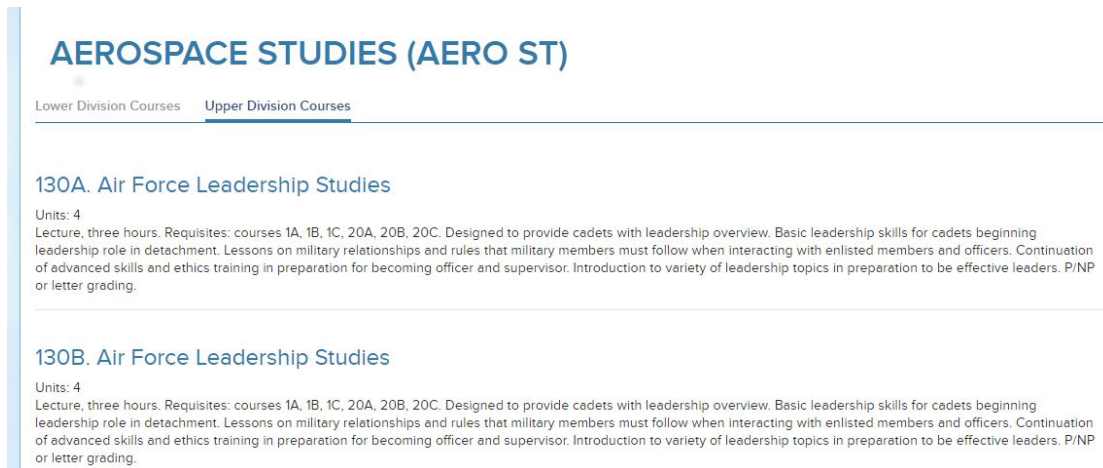
In this homework, you are asked to build a wrapper program that extracts data from the course descriptions of UCLA you crawled in the previous homework. Specifically, it should take a description page for a subject (example given below) and extract the following information from the page:

Subject: for example, "AEROSPACE STUDIES (AERO ST)".

Course number & title: for example, "130A. Air Force Leadership Studies"

Number of units: for example, "Units: 4".

Course description: for example, "Lecture, three hours. ... P/NP or letter grading."



**AEROSPACE STUDIES (AERO ST)**

Lower Division Courses   Upper Division Courses

---

**130A. Air Force Leadership Studies**

Units: 4  
Lecture, three hours. Requisites: courses 1A, 1B, 1C, 20A, 20B, 20C. Designed to provide cadets with leadership overview. Basic leadership skills for cadets beginning leadership role in detachment. Lessons on military relationships and rules that military members must follow when interacting with enlisted members and officers. Continuation of advanced skills and ethics training in preparation for becoming officer and supervisor. Introduction to variety of leadership topics in preparation to be effective leaders. P/NP or letter grading.

---

**130B. Air Force Leadership Studies**

Units: 4  
Lecture, three hours. Requisites: courses 1A, 1B, 1C, 20A, 20B, 20C. Designed to provide cadets with leadership overview. Basic leadership skills for cadets beginning leadership role in detachment. Lessons on military relationships and rules that military members must follow when interacting with enlisted members and officers. Continuation of advanced skills and ethics training in preparation for becoming officer and supervisor. Introduction to variety of leadership topics in preparation to be effective leaders. P/NP or letter grading.

For each page, create a JSON file to store the extract information as follows:

```
{
  "subject": "AEROSPACE STUDIES (AERO ST)",
  "courses": [{
    "course number & title": "130A. Air Force Leadership Studies",
    "Number of units": "Units: 4",
    "Course description": "Lecture, three hours, ..."
  },
  ...]
}
```

**Requirements:**

- Write your code in Python 3.
- Must use CSS selector or XPath for extraction.

**Submission:**

- **src:** folder contains your wrapper code.
- **results:** folder contains your extraction results (one JSON file per subject, which is named after the subject).